GTR Design Considerations

Overview

The GTR design parameters emerged from the SACGHS report on genetic testing. Ultimately, the goal is to collect sufficient information for each genetic test to document the basis and intended purpose(s) of the test, exactly what is tested and how, and the evidence that supports the test's Analytical Validity, Clinical Validity, and Clinical Utility. The first phase of GTR implementation is intended to balance these goals, the reporting burden on laboratories, existing resources, standards and practices, and the practicalities of measuring some of these attributes.

In general, GTR intends to collect information about a test directly from the group which developed and assayed the test; for laboratory-developed tests (LDT) this would be the laboratory that developed and performs the test. GTR generally will not register tests as belonging to a laboratory if the laboratory does not perform its own testing and simply sends tests out to other laboratories. However, the database attempts to accommodate some of the shades of gray that exist in this area: for example, where the assay, or one aspect of an assay, is sent out to another laboratory, but the interpretation is done by the offering laboratory. Another case might be a large laboratory network that contracts with laboratories outside its network to perform certain tests, but the contracting laboratory essentially performs as part of the network and only provides its testing services through such contractual arrangements. Another example would be an FDA-approved test, which may be performed in many laboratories, sometimes with laboratory-specific modifications; in this case, GTR would represent the FDA-approved test once, but then allow laboratories that use the test to "point to" the test's record for the supporting data on mechanism, validity, and utility instead of requiring the laboratory to submit the same information separately. Similarly, if a test is offered in several contexts by the same laboratory (for example both as a standalone test, and as a component of a panel or as a reflex test), it would be entered by the laboratory once, and then "pointed to" by the panel or reflex.

Basic Design

The GTR has 3 top-level data elements: Laboratories, People, and Tests. Most of the discussion in this document is on aspects of Tests, as this is where the most substantial portion of GTR resides, but some general issues of Laboratories and People are also discussed.

Laboratories

Because laboratories providing clinical testing services in the United States must be CLIA certified (or CLIA exempt), NCBI communicated with CMS about the possibility of using the CLIA database as a source of information for GTR. The primary information in the CLIA database that is relevant to GTR is laboratory demographic information and certifications. CMS was open to the idea and provided the CLIA database to NCBI. Unfortunately, information in the database was not sufficiently current to serve as a reliable source of up-to-date information on laboratories participating in GTR. As such, NCBI concluded

that laboratories would need to provide and update their contact and certification information directly in GTR. However, NCBI is attempting to make the laboratory fields compatible with the CLIA database schema so that in the future it may be possible to synchronize these two resources and eliminate the need for laboratories to enter the data in both places. GTR will also include tests from laboratories that are not CLIA certified (non-US laboratories and laboratories that only do research testing).

People

As with the Laboratories element, the primary data fields for the People element revolve around contact information and certifications. Names and contact information for individual people within a laboratory or a clinic may be important for several reasons: for example, some certifications are really certifications of an individual, not the laboratory, and therefore should follow the person if they were to change laboratories. In addition, there may be specific laboratory personnel who are the contacts only for particular tests, or who perform specific roles within the laboratory. GTR is designed so that the information about a person can be entered once, and then connected as appropriate to individual tests or specific roles within the laboratory. Much of this design is based on information collected and used by GeneTests now. The GTR design also includes provisions for personnel to choose whether their information should be publically shown in the database.

Tests

For purposes of GTR a Test is any separately orderable test offered by a laboratory. The field is not intended to define a general service area or disease for which a laboratory offers testing, but to explicitly identify each test that a physician could actually order for a patient. This simplifies the definition of "test" in the sense that it exactly corresponds to a laboratory's menu. It also means that GTR itself is not in the business of defining when a test offered by one laboratory is "the same" or "different" from that offered by another lab.

Each Test is automatically assigned a GTR accession number and version. This means every test can be explicitly cited by other resources, publications, or guidelines in an unambiguous way. Since the tests are versioned, and older versions retained (the latest version is shown by default), it is possible to retrieve the exact test information as it appeared at the time such outside documents were written, as well as to readily see that later versions now exist. This provides a mechanism for scientists, clinicians, and other groups outside GTR staff to publish their own assessments of Tests, such as which Tests are the "same" or "similar," or which may be more appropriate for specific uses. This is not a GTR activity per se, as these judgments are matters of significant scientific or clinical expertise, but GTR accession/version numbers will facilitate such activities and make these associations much more explicit.

Analytes

GTR attempts to be as objective and explicit about the details of testing as possible, while recognizing that a laboratory may consider some information about the test proprietary and that laboratories will likely differ in the level of detail they are willing to share. Thus, GTR supports the provision of considerable detail on the basis of a test but allows the laboratory discretion in the level of detail it provides.

For tests targeted at the genome, at a minimum GTR would like to receive something which resolves to coordinates on the human genome. Simply naming a human gene does not accomplish this, since information about the extent or number of exons of a gene continues to change with new knowledge, and because nomenclature may not be sufficiently consistent or objectively verifiable. Independent of how laboratories may measure the relevant sequence regions of a patient's genomic DNA, they will still typically call mutations over specific pieces of reference DNA. So, as a minimum, GTR would request a laboratory to supply those reference sequence regions as FASTA files. These reference sequences are usually public sequences and often do not correspond to the full or precise regions actually sequenced, so no proprietary information would be revealed. Nevertheless, providing this information is important: it allows GTR to objectively verify which regions of the genome are being assayed, to provide the submitter with confirmation that no errors occurred in submission and that GTR correctly represents their test, and it is a simple way to provide exact coordinates on the genome without error prone-entry of large integers by the submitter. As an additional benefit, GTR would be able to proactively report back to submitters when new information becomes available that is relevant to their test (e.g., a new alternate transcript exon that may not be fully covered by their current testing strategy, or new SNPs discovered in the region).

Similarly, GTR would ask for explicit lists of mutations with clinical significance being assayed by the tester. It is understood that this is a moving target, but establishing such lists, along with knowing the genome regions assayed, is another step toward developing an informative two-way flow of information between the research community and clinical laboratories.

Analytical and Clinical Validity

One of the stated purposes of GTR is to allow the testing laboratories to explicitly state the Analytical and Clinical Validity of their tests. In practice, however, this can be problematic to do in detail. Specific components of Analytical or Clinical Validity, such as population-based measures of sensitivity or specificity, may not be possible to obtain, for example in the case of rare diseases. Many aspects of Clinical Validity are not determined by the laboratory at all, but are obtained from the research literature and thus are generic to testing for a particular disease and not specific to a particular test.

In general, GTR has chosen to represent these fields as a block of text, supported by references, rather than having submitters fill in a large number of specific fields. If available, guidance will be provided for the types of elements that should be included, but the open structure will allow the laboratory the flexibility to describe these aspects of their test appropriately for many different situations, while still supporting their assertions with published sources. This structure also reduces the reporting burden, since many laboratories already have very good documents on their web sites and could cut and paste text that has already been produced instead having to fit the information into new fields.

Some additional fields attempt to add objective and easily stated indicators of analytical validity and quality control assessment, such as whether the test undergoes proficiency testing and if so, what sort.

GTR recognizes that many clinical laboratories already gather evidence of clinical validity as part of their certification or licensure requirements. For example, New York State requires evidence of clinical validity

for licensure, and the College of American Pathologists (CAP) – one of the key groups that performs CLIA certification – includes items about clinical validity on its "Molecular Pathology Checklist," a list of items that a CAP inspector can expect a laboratory to have available to show for any of their tests. Proposed fields for GTR (see the Proposed GTR Field Definitions document) include elements that correspond to some of these CAP Checklist items. Because the items are on the Checklist, the assumption is that any CAP-certified laboratory should have the information conveniently available to populate these fields. GTR's intent here, as in other areas, is to try to minimize reporting burden by conforming to existing sources and requirements.

Clinical Utility

There has been substantial input from stakeholders and advisors that Clinical Utility may not be an appropriate field to collect for GTR. Among the reasons given are that Clinical Utility is the most fluid attribute of a test, that it may vary considerably based on the clinical context of an individual patient, and that it is not necessarily a property of the test itself in isolation. However, as with Clinical Validity, we defer to current practice, and ask laboratories to provide information that corresponds to items such as CAP Checklist MOL.30670 "Clinical Indication/Clinical Utility" (which asks the laboratory to provide that section of their procedure manual that describes "clinical utility of the test in patient management, with pertinent literature references").

Test Codes

There have been different views about the utility of including codes such as ICD9, ICD10, or CPT codes, which are diagnostic and billing codes. The argument against including them is that these codes do not code for the test, but for the diagnostic or economic context of the test, and their appropriate use depends on the clinical and insurance context, not necessarily on the test itself. The argument for including them is that some people feel they are useful search terms (particularly the diagnostic codes), they may facilitate quick location and ordering of a test, and they may be helpful to users looking for a less common test. Experience likely will be the best guide as to whether these codes are useful, so GTR will allow laboratories to deposit the codes if they wish to do so, and will evaluate their utility over time.

Specimen Source

GTR is attempting to strike a middle ground on this element. Basic information about specimen collection (such as "whole blood" or "saliva") may be very useful to users to limit search results or subset tests, and thus are included as GTR data elements. However, very detailed sample preparation or shipping requirements are probably beyond the scope of GTR and may be best handled as a link to the testing laboratory website since they may change frequently, and GTR users are not likely to want to search by these descriptors.

Bulk Data Exchange

GTR will support interactive web-based forms for data entry and will provide user friendly web pages for searching and viewing the data.

In addition, GTR also will support the ability to download the contents of the database, in XML, for incorporation into other information systems, as well as mechanisms for bulk deposit by large laboratories that may have existing information systems and where the investment in software development would counterbalance a larger investment in staff time to update hundreds of tests through the web forms.

Most users and submitters will only see the web forms and will not need to understand this level of detail. However, the web forms will reflect the same rules and design (through what they present and how they interact) as those explicitly expressed in the technical XML specification. In the "Proposed GTR Field Definitions" document there are a few fields, mostly identifiers, which would only be used by bulk submitters, and it is expected that this level of detail will be of interest only to a small number of implementers.

XML Exchange

The XML design (XSD) has three types: [Laboratory + List of People], [List of Tests] and [ClinVar]. Tests may be added, deleted, or updated more frequently than Laboratories or People, and thus is a separate section. The separate Clinvar section allows clinical validity data to be updated or exchanged separately as more information is gathered. It also makes it possible for a Test to simply point to a pre-existing ClinVar record for well-established mutations, without needing to re-enter this information. This XML will be the base for both bulk update and for bulk download. Additional information is necessary for the transaction itself (data submission or data exchange), but these are the core content elements.

GeneTests

There is no supported bulk download for GeneTests; however, since NCBI also supports the GeneTests database, we have designed a database-to-database transfer to load GTR with data from the GeneTests Laboratory Directory in an ongoing way. This will facilitate rapid transfer of existing information for labs in the GeneTests Laboratory Directory that wish to participate in GTR, and will provide them with a starting point based on what they already submitted to GeneTests. The GTR design will also include mechanisms to ensure that a search of GTR will capture any information in the GeneTests Laboratory Directory that has not yet been transferred to GTR during the transitional period when both resources are in operation.

HL7 e-DOS

There is an HL7 standard for an electronic Directory of Services (e-DOS) which is intended for a laboratory to provide an HL7 message to a Hospital information system describing the tests the laboratory offers in a computer-readable form. In principle, this is very similar to the XML bulk upload (or download) for GTR. However, the purposes are somewhat different in that GTR captures much more detail on the basis and interpretation of the test, and much less detail on specimen collection and billing information. Nonetheless, GTR is working with the HL7 e-DOS group to maximize compatibility in the data design, at a minimum, and to ease implementation of data management for groups supporting both. The best case would be interchangeable exchanges, but this may not be practically attainable.